DOCUMENT RESUME

ED 237 532 TM 830 782

AUTHOR Klein, Stephen P.; Bolus, Roger
TITLE The Effect of Item Sequence on Bar Examination Scores.
INSTITUTION Rand Corp., Santa Monica, Calif.
REPORT NO Rand-P-6857
PUB DATE Apr 83
NOTE 8p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Montreal, Quebec, April 12-14, 1983).
AVAILABLE FROM Rand Corp., 1700 Main Street, Santa Monica, CA 90406 (P-6857-$4.00).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS Adults; *Cheating; Cost Effectiveness; Item Analysis; *Multiple Choice Tests; Scores; Test Construction; *Test Format; *Testing Problems; Testing Programs; *Test Items
IDENTIFIERS *Item Position (Tests); *Multistate Bar Examination; National Conference Of Bar Examiners

ABSTRACT

A solution to reduce the likelihood of one examinee copying another's answers on large scale tests that require all examinees to answer the same set of questions is to use multiple test forms that differ in terms of item ordering. This study was conducted to determine whether varying the sequence in which blocks of items were presented to examinees would affect test and/or item characteristics. Sixty items drawn from four content areas of a secure version of the Multistate Bar Examination (MBE) were constructed into four test forms and administered to 2940 bar applicants randomly divided into four groups. The testing design provided two independent tests of sequence effects under a 55 minute time limit and a 90 minute time limit. The four groups had almost identical means and standard deviations on the full 200 item MBE. The findings indicate that variations in the order in which blocks of MBE items were asked had little or no effect upon test or item statistics. This was true under the regular time per item as well as under almost total power conditions. The use of such forms appears to be a psychometrically sound and cost effective method for discouraging cheating in those testing programs that face the same policy constraints as are encountered on bar examinations. (PN)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED237532

THE EFFECT OF ITEM SEQUENCE ON BAR EXAMINATION SCORES

Stephen P. Klein

Roger Bolus

April 1983

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

T. Cookwell

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM820 782

P-6857

The Rand Paper Series

Papers are issued by The Rand Corporation as a service to its professional staff. Their purpose is to facilitate the exchange of ideas among those who share the author's research interests; Papers are not reports prepared in fulfillment of Rand's contracts or grants. Views expressed in a Paper are the author's own, and are not necessarily shared by Rand or its research sponsors.

The Rand Corporation
Santa Monica, California 90406

THE EFFECT OF ITEM SEQUENCE ON BAR EXAMINATION SCORES

Stephen P. Klein, The Rand Corporation
Roger Bolus, GANSK & Associates

Large scale testing programs can reduce the likelihood of one examinee copying another's answers by: having sufficient distance between seats, having adequate proctoring, varying answer sheet format, and using multiple test forms. The use of multiple forms usually involves having one form contain one set of items and the other forms contain different sets of items. In other words, at a given administration of the examination, all examinees do not answer the same questions. Although this strategy may be sound in terms of psychometric standards, it may be inconsistent with the policies of the organization sponsoring the testing program. For instance, the National Conference of Bar Examiners requires that all examinees taking the Multistate Bar Examination (MBE) on one of its biannual administrations answer the same set of questions.

The MBE is a 200 item multiple choice test that is taken by about 55,000 applicants to the bar each year. In many states, there is often substantially less than adequate seating distance among examinees. This situation has led to several incidents of cheating. And, cheating on a bar examination is especially serious because it is a moral character violation that may prohibit an examinee from practicing law even if he/she retakes and passes the examination.

One solution to the foregoing problem is to use multiple test forms that differ in terms of the order in which the items appear. This strategy is consistent with the policy of having all examinees answer the same items

---

This paper was presented at the annual meetings of the National Council on Measurement in Education, Montreal, April 12, 1983. The views expressed in this paper are the authors and do not necessarily represent those of The Rand Corporation, GANSK & Associates, or the National Conference of Bar Examiners.

and it would substantially reduce the opportunity for cheating. There are two major concerns with this approach: (1) some sequences may be easier than others thereby giving some examinees an unfair advantage and (2) it might change the characteristics of items used for equating tests across administrations.

There is a no data on whether essentially random variations in item sequence would change the psychometric properties of a test or its items. Almost all the literature on item order effects comes from studies with high school or college students. These studies have investigated systematic rather than random variations in item sequence (such as from easy to hard versus hard to easy) and/or the effects of mixing versus separating item types or content (such as quantitative and verbal items). These studies are therefore not especially relevant to the MBE and many other large post secondary testing programs.

PURPOSE

The present study was conducted to determine whether varying the sequence in which blocks of items were presented to examinees would affect test and/or item characteristics. There were two reasons for studying the effects of varying blocks rather than individual items: (1) many tests, including the MBE, have several items tied to a common passage and (2) it would be less expensive to print and score multiple forms if variation was limited to item blocks.

SAMPLE

The sample for the study consisted of 2940 applicants to the bar in a large western state. These applicants were encouraged to participate and do well in the study because a high score would improve their chances of passing the MBE and essay portions of their state's bar examination.

INSTRUMENTS

The study used 60 items that were drawn from 4 content areas. These items had appeared on previous but still secure versions of the MBE. The items were divided into two sets, A and B. Each set contained 30 items.

Two versions of each set were constructed. Thus, there were a total of four forms: A-1, A-2, B-1, and B-2. The first 10 items on form A-1 were the same as the last 10 on A-2 while the last 10 on A-1 were the same as the first 10 on A-2. Forms B-1 and B-2 followed this same XYZ and ZYX pattern. Table 1 shows how items were allocated to test forms.

Table 1

ASSIGNMENT OF ITEMS TO FORMS

Test Form	Sequence of items within form
A-1	1-10, 11-20, 21-30
A-2	21-30, 11-20, 1-10
B-1	31-40, 41-50, 51-60
B-2	51-60, 41-50, 31-40

PROCEDURES

Applicants were assigned randomly to 4 groups. Groups 1 and 3 took an A form under a 55 minute time limit (which is the normally allowed time per item) and then a B form under almost total power conditions (a 90 minute time limit). Groups 2 and 4 took a B form under a 55 minute time limit and then an A form under a 90 minute time limit. This design, which appears in Table 2, provides two independent tests of sequence effects under the 55 minute time limit (Groups 1 vs 3 on set A and 2 vs 4 on set B) and two independent tests under the 90 minute time limit (Groups 1 vs 3 on set B and 2 vs 4 on set A).

Table 2

ASSIGNMENT OF FORMS TO GROUPS

Time limit	Group 1	2	3	4
55 minutes	A-1	B-1	A-2	B-2
90 minutes	B-1	A-1	B-2	A-2

A = items 1 to 30, B = 31 to 60

RESULTS

The four groups had almost identical means and standard deviations on the full 200 item MBE (means ranged from 428.8 to 430.4).

The average mean score, standard deviation, and coefficient alpha on a set of 30 items taken under the 55 minute time limit were 20.32, 3.96, and .645 respectively. The corresponding values under the 90 minute time limit were 21.19, 3.88, and .650. Table 3 shows the differences in these three statistics between groups under each time limit that were due to the variation in item order. None of the small observed differences in test statistics attributable to item sequence even approached statistical or practical significance.

Table 3

DIFFERENCES IN TEST STATISTICS DUE TO ITEM SEQUENCE

Time limit	Groups compared	Item set	Mean score	Standard deviation	Coeff alpha
55	1 vs 3	A	.01	.05	.01
	2 vs 4	B	.12	.04	.01
90	1 vs 3	B	.16	.28	.06
	2 vs 4	A	.15	.26	.04

A 55 minute time limit for 30 items is consistent with the amount of time per item on the MBE.

The total scores on a form correlated about .70 with the scores on the regular 200 item MBE that was taken on the following day. Variations in item sequence did not significantly affect this relationship. For instance, Groups 1 and 3 had r's of .71 and .69, respectively, under the 55 minute limit. Both groups had an r of .74 under the 90 minute limit. Variations in sequence also did not affect relationships with scores on the essay portion of the regular bar examination.

Under the 55 minute time limit, the means on the 30 items on Form A-1 correlated .98 with their means on Form A-2. The correlation was .99 under the 90 minute time limit. The corresponding values with the B forms were .98 and .96. In short, item difficulties as well as total test statistics were insensitive to variations in item sequence. Correlations among z transformed item biserials averaged .78 under the 55 minute limit and .73 under the 90 minute limit, however, there was much less variation among the biserials on a form than there was among that form's item difficulties.

DISCUSSION AND CONCLUSIONS

The foregoing findings indicate that variations in the order in which blocks of MBE items were asked had little or no effect upon test or item statistics. This was true under the regular time per item as well as under almost total power conditions. Thus, neither an examinee's score or the process of equating tests across administrations would be affected by the use of multiple forms in which the sequence of items was varied. The use of such forms therefore appears to be a psychometrically sound and cost effective method for discouraging cheating in those testing programs that face the same policy constraints as are encountered on bar examinations.